

樊原廷

(+86) 183-9112-6180 | fan0yuanting@gmail.com

AIGC 算法研究员 · 腾讯优图实验室

retofan233.github.io | retofan-blog.pages.dev | [Google Scholar](https://scholar.google.com/citations?user=...)

西安交通大学 软件工程硕士 | 2022.09 - 2025.06 长安大学 软件工程学士 | 2018.09 - 2022.06

工作经历

聚焦 3D 纹理生成、工业质检缺陷合成与低成本 AI 漫剧生成；独立主导数据、训练、评测到上线全链路，获 2026H1 Outstanding。

 腾讯科技（深圳）有限公司 · 优图实验室

Researcher | 2025.06 - 至今

Youtu3D (YT3D) | 单/多参考图纹理生成

- 背景与目标：**两个月内独立主导 Youtu3D 单/多参考图纹理生成模型的数据、训练、评测与上线全链路，目标是在细小文字生成、多图参考遵循、以及真实世界物理还原层面做到业内顶级水平。
- 数据工程：**自研 Blender 分布式 PBR 渲染管线，产出 150 万+ 多视角高质量渲染数据；引入 OCR 模型对文字场景数据进行增广，并引入透明 / 镂空资产合成 Blender 管线缓解长尾数据分布。
- 模型与训练：**完成多阶段分辨率 scaling 训练；自研 Klein 4B/9B 多参考 albedo 框架，注入视角位姿，用视角相关注意力让目标 albedo 只与同朝向参考图交互，并加光照一致性约束，减轻多视图色差与接缝。
- 透明 / 镂空 (MRT)：**主导业内首个支持透明 / 镂空 3D 资产生成的纹理模型：以 MRT 将金属度、粗糙度与透明度拆到多通道材质图联合生成，区分镂空与透射并回写 GLB，打通训练到导出全链路。
- 规模化迭代：**基于 300+ 卡异构 GPU 集群完成多轮 SFT / DPO / GRPO；建立 50+ 轮、超 200 人的 GSB 评测闭环。
- 效果：**模型在细小文字生成、多图参考遵循、以及真实世界物理还原层面全面领先；客观指标领先业内最好模型超过 30%；单/多参考及透明 / 镂空生成已上线腾讯云国内 / 国际站，客户反馈明显优于现网最新的混元和 Tripo 模型；获 2026H1 Outstanding 绩效。

UniDG | 工业质检通用缺陷合成基模 [arXiv:2604.08915](https://arxiv.org/abs/2604.08915)

- 背景与目标：**工业质检异常样本稀缺，既有 few-shot 缺陷生成易过拟合特定类别；目标是构建无需按类微调、可同时支持参考图迁移与文本指令编辑的通用缺陷合成基模。
- 数据建设：**构建 UDG 数据集，含 30 万组 normal-abnormal-mask-caption 四元组，覆盖 50 个公开数据集、269 类原始缺陷（归并为 28 类），规模约为既有最大缺陷生成数据集的 23 倍；以 Inpainting / Captioner / Verifier 多 Agent 管线完成配对与质检。
- 模型设计：**基于 FLUX.1-Fill-dev，提出 Defect-Context Editing（自适应缺陷裁剪 + diptych 结构化输入），并通过 MM-DiT 多模态注意力融合参考条件与目标区域。
- 训练策略：**两阶段训练 Diversity-SFT + Consistency-RFT（Flow-GRPO + 双奖励模型），在多样性与参考一致性之间取得平衡，支持零样本跨类缺陷迁移。
- 效果与落地：**MVTec-AD / VisA 上合成质量与下游检测定位全面领先；多类分割 FG-mIoU 相对 AnoGen 提升 20.5%；落地腾讯云 TIAOI 并服务多个 B 端客户；论文投稿 NeurIPS 2026。

AI 漫剧 / 短剧生成平台 | 小说→分镜→成片

- 背景：**围绕低成本 AI 漫剧 / 短剧生成，构建从平台到模型的完整技术链路，服务中心多部门及 B 端客户。
- 平台链路：**一键式短剧生成（小说→自动分镜剧本→图像 / 视频生成→FFmpeg 在线剪辑合成）；Python 后端 + Vue 前端 + OpenAI 兼容模型服务，整合多模态理解、图像视频生成、视频超分与插帧。
- 工程能力：**统一模型网关与任务队列，支持分镜级重试、素材缓存与成片导出；将开源理解 / 生成模型封装为可插拔服务，降低业务接入成本。

音画同出视频生成 | 多 Speaker DPO

- 数据与奖励：**针对多 Speaker 场景构建 Speaker 级正负偏好数据；引入声纹-人脸 embedding 对齐损失作为奖励模型，约束口型、身份与音色一致性。
- 训练：**基于 LTX-2.3-22B 进行多尺度 DPO（全局质量 / Speaker 级 / 时段级），同时优化整体观感与角色级音画对齐。

高清 Any2Any 换脸 | Wan2.2 + 4-Step 蒸馏

- 数据：**基于 Ray Data + LanceDB 构建大规模 4K 人脸视频切片（短边 >768px），支持高效检索与可复现训练。
- 训练与加速：**基于 Stand-In Wan2.2-T2V-A14B 完成多阶段渐进训练（256→768px）+ 混合采样 + LightX2V 4-Step 蒸馏；内部评测在参考遵循与人脸动态上大幅领先既有 SOTA。

实习经历

 腾讯科技（深圳）有限公司 · 优图实验室

Research Intern | 2024.04 - 2024.10

工业质检多模态大模型 | 工具调用 + 异常检测

- 全流程主导：**完整负责工业质检多模态大模型的数据、训练与部署；制作高质量指令微调与工具调用数据，经 LLaMA-Factory 完成 SFT 与人类反馈学习。
- 工具链：**结合开源学术 / 工业异常检测数据训练后端工具模型，含 MM-GroundingDINO（开集检测）与 WinCLIP（异常检测）；经 vLLM 部署，并以 Gradio / Streamlit 提供稳定交互接口。
- 研究沉淀：**深入细粒度多模态大模型、开集检测与视觉交互式检测，后续工作延伸为 PR 2026 一作论文 FineGrainedAD。

科研成果

NeurIPS 2026 在投 / 一作 Large-Scale Universal Defect Generation [arXiv:2604.08915](#)

- 问题：工业 / 医学等场景异常样本稀缺，few-shot 缺陷生成易过拟合类别，难以跨物体迁移。
- 贡献：构建 UDG（30 万组 normal-abnormal-mask-caption 四元组，覆盖 50 个数据集、269 类缺陷）；提出 UniDG，基于 FLUX.1-Fill-dev 与 Defect-Context Editing，无需按类微调即可做参考图迁移与文本指令编辑。
- 结果：MVTec-AD / VisA 上合成质量与下游检测定位全面领先；多类分割 FG-mIoU 相对 AnoGen 提升 20.5%；落地腾讯云 TIAOI。

Pattern Recognition 2026 / 一作 Towards Fine-Grained Vision-Language Alignment for Few-Shot Anomaly Detection

- 问题：既有 VLM 小样本异常检测方法用图像级文本去匹配 patch token，语义错位导致定位偏弱。
- 贡献：提出 MFSC 自动构建多层级细粒度语义描述，以及 FineGrainedAD（MLLP 可学习提示 + MLSA 多级对齐），让提示词与视觉区域对齐。
- 结果：在 MVTec-AD 与 VisA 的 few-shot 设定下整体定位性能优于既有方法。

ECCV 2024 / 一作 AdaDiffSR: Adaptive Region-aware Dynamic Acceleration Diffusion Model for Real-World Image Super-Resolution

- 问题：扩散超分对整图使用统一去噪步数，在不可见细节上浪费计算。
- 贡献：用多指标潜空间信息增益（MMLE）动态选择区域时间步，并以渐进特征注入（PFJ）保持保真度与真实感。
- 结果：在合成与真实数据上达到与当时 SOTA 相当的重建质量，计算量约为 StableSR 的 60%。

IEEE TCSVT 2025 / 一作 InstanceSR: Efficient Reconstructing Small Object with Differential Instance-level Super-Resolution

- 问题：常规超分优化整图观感，对小目标聚集区域分配不足，限制后续检测。
- 贡献：按实例分布将图像编码为不同难度 token，用多出口解码器对前景实例消耗更多算力、对背景少算。
- 结果：在含小目标的超分基准上优于既有方法，并稳定提升后续小目标检测精度。

AAAI 2024 / 学生二作 Decoupling Degradations with Recurrent Network for Video Restoration in Under-Display Camera

- 问题：屏下相机视频同时存在眩光、雾状散射等随时间变化的退化，且需保持时序一致性。
- 贡献：提出 D2RNet，用解耦注意力分别处理 flare / haze；构建基于实测 PSF 的大规模 UDC HDR 视频基准。

竞赛、专利与专业能力

- 竞赛与专利：2023 中国研究生电子设计大赛西北赛区一等奖；2021 中国软件杯国家三等奖；多项专利受理。
- 算法：熟悉扩散 / Flow Matching、SFT / DPO / Flow-GRPO、3D 纹理生成、多模态大模型、异常检测与缺陷合成、超分辨率。
- 工程：熟练使用 Python、PyTorch、LLaMA-Factory、vLLM、Ray、Blender PBR 渲染、Docker、Linux、Git；能独立完成从数据到训练、评测、服务化上线的全流程（One Person Team）。
- 方向：目前关注 Agent 应用、低成本 AI 漫剧 / 短剧生成、统一生成与理解模型、世界模型等前沿问题。欢迎邮件联系。